MAS223 Statistical Modelling and Inference Chapter 2: Multivariate distributions

Dr Nic Freeman

School of Mathematics and Statistics, University of Sheffield

Autumn Semester, 2015

Introduction

If we observe several (k) numerical quantities in the same experiment, for example for lung cancer patients

$$\begin{aligned} X_1 &= \text{age} \\ X_2 &= \text{size of tumour} \\ X_3 &= \text{smoking level} \\ X_4 &= \text{socio-economic group} \end{aligned}$$

then we have a **multivariate** random variable or random **vector**

$$\mathbf{X} = (X_1, X_2, \ldots, X_k)^T.$$

Definitions

Formally, $\mathbf{X}$ is a mapping from the sample space S into k-dimensional space $\mathbb{R}^k$.

If all $X_1, X_2 \ldots, X_k$ are continuous random variables $\mathbf{X}$ is said to be a **continuous random vector.**

If all $X_1, X_2, \ldots, X_k$ are discrete random variables $\mathbf{X}$ is said to be a **discrete random vector.**

Otherwise, $\mathbf{X}$ is neither continuous nor discrete.

Definitions

You will already have met discrete random vectors in MAS113; in this course we will:

- Extend the ideas there to discuss continuous random vectors.
- Introduce the important case of vectors with multivariate normal distributions.
- Look at transformations of multivariate distributions.

Often we will concentrate on the **bivariate** case $k = 2$ and denote the random variables by X and Y.

Joint distribution and density functions

If X and Y are any two jointly distributed random variables then we can define their **joint distribution function** as

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

for all real x and y.

In principle, if we know the value of this function for all x, y then we can evaluate all other probabilities involving X and Y.

Joint distribution and density functions

If $F_{X,Y}$ is sufficiently smooth to possess the partial derivative

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

(except possibly, for example, on the boundary of a region) then $f_{X,Y}$ is called the **joint probability density function (p.d.f.)** of X and Y.

It is analogous to the p.d.f. of a univariate distribution, and may be thought of as measuring the "probability per unit area" at each point (x, y) in the plane.

Joint distribution and density functions

To calculate the probability that the pair (X, Y) lies in some region D of the plane we can integrate $f_{X,Y}$ over D:

$$P((X, Y) \in D) = \int\int_D f_{X,Y}(x, y)\, dx\, dy.$$

Pictorially, if we plot the surface $z = f_{X,Y}(x, y)$ in three dimensions then this probability is the **volume** between this surface and the plane $z = 0$ determined by the set D in the (x, y) plane.

Joint distribution and density functions

Note that if we choose $D = (-\infty, x] \times (-\infty, y]$ in the above then we get

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) = P((X,Y) \in D) \\ &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\, du\, dv. \end{aligned}$$

Since evaluating probabilities involves double integration, often over a bounded region, it is important to get the **limits** of integration right.

For example if we integrate with respect to y first then we need to ascertain the limits of y for each fixed x.

Joint distribution and density functions

Example 10: Joint probability density function; calculating probabilities

Joint distribution and density functions

For any bivariate p.d.f., we have

1. $f_{X,Y}(x,y) \geq 0$ for all x, y;
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$.

Marginal distribution

Where we have a multivariate random variable $(X_1, X_2, \ldots, X_k)$, the **marginal distribution** of a component X_i is simply the distribution of X_i considered as a univariate random variable.

Marginal distribution

We can find the marginal distribution of X by "integrating out" the other variables.

In the bivariate case (X, Y), the marginal p.d.f. of X is found by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

("integrate y out").

If $f_{X,Y}(x, y)$ is only positive on a restricted region then the effective limits of integration may depend on the value of x.

Conditional distribution

The conditional p.d.f. of Y given $X = x$ is given by the ratio of the joint p.d.f. and the marginal p.d.f. of the variable being conditioned on:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

provided $f_X(x) > 0$.

Example

Example 11: Marginal and conditional distributions

Covariance

Let $\mu_X = E(X), \mu_Y = E(Y)$. The **covariance** $\mathrm{Cov}(X, Y)$ is defined as

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

Here $E(XY)$ must be calculated as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_{X,Y}(x, y)\, dx\, dy.$$

Correlation

The **correlation coefficient** $\rho(X, Y)$ is then defined as

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{(\mathrm{Var}(X)\,\mathrm{Var}(Y))}}.$$

These measure the extent to which X and Y vary together.

[If you have a set of data which are a random sample from the distribution of (X, Y), Pearson's sample correlation coefficient, which some of you may have seen, is an estimator of $\rho(X, Y)$.]

Example

Example 12: Covariance and correlation

Independence

If we have independent random variables X and Y with p.d.f.s $f_X(x)$ and $f_Y(y)$ respectively, then the random vector (X, Y) has joint p.d.f. given by

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Therefore, if X and Y are independent, we can find functions $g(x)$ and $h(y)$ such that $f_{X,Y}$ factorises in the form

$$f_{X,Y}(x, y) = g(x)h(y)$$

for all x, y.

Independence

The reverse implication is also true. If we have

$$f_{X,Y}(x,y) = g(x)h(y)$$

for two p.d.f.'s g and h then this implies that X and Y are independent. In fact, then g and h are necessarily the marginal p.d.f.'s of X and Y since for instance

$$f_X(x) = \int_{-\infty}^{\infty} g(x)h(y)\, dy = g(x)\int_{-\infty}^{\infty} h(y)\, dy = g(x).$$

Example

Example 13: Independence

Conditional expectation

The **conditional expectation** of Y given that X takes the value x can be defined as the expectation of a random variable with the conditional distribution:

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy.$$

Note that this is a function of x, $g(x)$ say. As X itself is a random variable, so is $g(X)$. We define

$$E(Y|X) = g(X),$$

which is a random variable; if we know the value of X is x then we know the value of $E(Y|X)$ is equal to $E(Y|X = x)$.

Example

Example 14: Calculating conditional expectation

Conditional variance

We may similarly define the conditional variance as the variance of the conditional distribution of Y given X, namely

$$\mathrm{Var}(Y|X) = E\{(Y - E(Y|X))^2|X\}.$$

We can also define conditional covariances: if X, Y and Z are random variables, then we can define the conditional covariance of X and Y, given Z, as:

$$\mathrm{Cov}(X, Y|Z) = E(XY|Z) - E(X|Z)E(Y|Z).$$

Properties of conditional expectations

1. $E(Y) = E\{E(Y|X)\}$.
2. $\mathrm{Var}(Y) = E\{\mathrm{Var}(Y|X)\} + \mathrm{Var}\{E(Y|X)\}$.
3. $\mathrm{Cov}(X, Y) =$ $E\{\mathrm{Cov}(X, Y|Z)\} + \mathrm{Cov}\{E(X|Z), E(Y|Z)\}$.
4. $E(Yf(X)|X) = f(X)E(Y|X)$

Properties 1 and 2 are useful when the best way of finding the mean and variance of Y is by conditioning on X. Property 4 says that any factor which is a function of X only may be taken outside the expectation.

Example 15: Proof of Property 1

Example 16: Calculation of expectation and variance by conditioning